Discrete Applied Mathematics 10 (1985) 197-201
North-Holland

ON SOME CONDITIONING RESULTS IN THE PROBABILISTIC ANALYSIS OF ALGORITHMS

Colin McDIARMID
Institute of Economics and Statistics, Oxford, U.K.

Received 15 January 1984

A simple combinatorial approach is given for handling certain conditioning problems that arise in the probabilistic analysis of graph algorithms.

A key step in the probabilistic analysis of combinatorial algorithms is often that of establishing that certain conditioning introduced by the operation so far of the algorithm either helps or at least does not hurt too much. In this note we consider a conditioning problem that arises from searching adjacency lists, and which occurs for example in the analysis of algorithms for finding Hamiltonian cycles, perfect matchings, connected components or blocks [1], [2]. We introduce a natural method of generating the random input which reduces the problem to a simple combinatorial one.

We are interested in the probabilistic analysis of algorithms which operate on certain random subgraphs R of some fixed (finite, simple) graph G. This graph G may correspond to a part of another graph, for example a complete graph K_n, not explored by a previous phase of the algorithm. We suppose that the random graph R is presented as a family of ordered adjacency lists. At each step, depending on what has happened so far, the algorithm either will terminate or will choose some vertex v and look at the next entry in the adjacency list for v. If it *finds* vertex w then the edge $\{v, w\}$ has been *selected.*

For each possible random graph R we suppose that each possible ordering of the adjacency lists is equally likely. Also we suppose that each subgraph of G with the same number of edges is equally likely (see Remark (ii) below). For each vertex v of G let D_v denote its degree in the random graph R. Our main result is the following.

Proposition 1. *Let v, x, y be distinct vertices of G. Suppose that at some step the algorithm is about to look at the next entry in the adjacency list for vertex v, that both the edges $\{v,x\}$ and $\{v,y\}$ are in G and have not been selected, and that the adjacency lists for vertices x, y have been examined r_x, r_y times respectively (without finding an end-marker). Then for any integers $d_x \geq 0$ and $d_y \geq 1$ for which*

$$P\{D_x \geq d_x, D_y \leq d_y\} > 0$$

0166-218X/85/$3.30

we have

$$P\{\text{find}\, y \mid D_x \geq d_x, D_y \leq d_y\}$$
$$\leq P\{\text{find}\, x \mid D_x \geq d_x, D_y \leq d_y\} \max\left\{1, \left(1 - \frac{r_y}{d_y}\right) \Big/ \left(1 - \frac{r_x}{d_x+1}\right)\right\}.$$

Before we prove this result let us note two immediate corollaries and make some remarks.

Corollary 2. *If* $P\{D_x \geq d_x\} > 0$, *then*

$$P\{\text{find}\, x \mid D_x \geq d_x\} \geq P\{\text{find}\, y \mid D_x \geq d_x\}\left(1 - \frac{r_x}{d_x+1}\right).$$

Corollary 3. *If* $r_x = 0$, *then*

$$P\{\text{find}\, x\} \geq P\{\text{find}\, y\}.$$

Remarks. (i) This last result may also be deduced using the approach in [1]. It yields lemmas 1 and 5 of [2]: we take G to have the edges of the complete graph K_n not found in the first phase of the algorithm, and make all subgraphs of G with the appropriate number of edges equally likely.

(ii) It will be seen that our proof allows a more general distribution for the random graph R, namely that for all subsets F of edges of G not containing $\{v, x\}$ or $\{v, y\}$

$$P\{R = F \cup \{v, y\}\} \leq P\{R = F \cup \{v, x\}\}.$$

Here we are identifying the random subgraph R of G with its edge-set in the obvious way. This more general assumption allows us to handle a random graph to which certain extra 'good' edges have been added. For example, in lemma 5 of [2] we might wish to add to our random graph as above all the edges found in the first phase of the algorithm which have one end in the giant block and one end outside it. It is not hard to see that if G is connected we may not relax further our assumption on the distribution of random graph R.

(iii) Corresponding results for random directed graphs – see lemma 3 of [2] – may be deduced from the above by splitting each vertex of the underlying directed graph into an 'in-vertex' and an 'out-vertex' and then undirecting edges.

(iv) Our treatment here is related to that of the 'Almost Equiprobability (AEP) Lemma' of [1]. Suppose that all graphs with n labelled vertices and $N = N(n)$ edges are equally likely. Suppose that at some stage, for each vertex we have selected only $o(n)$ incident edges and we have examined at most $r \log n$ entries from its adjacency list. We are about to look at the next entry in the adjacency list for some vertex v. Let x be an eligible vertex (that is, $x \neq v$ and the edge $\{v, x\}$ has not been selected), and let S_x be the event that x is the next eligible vertex in the adjacency list for vertex v. Let $Z(K)$ be the event that all vertex degrees are at least Klogn. Then

$$\frac{1}{(1+\varepsilon)n} < P(S_x \mid Z(K)) < \frac{1+\varepsilon}{n}$$

if $N \geq (2K+4)n \log n$ and n is sufficiently large, where

$$(1+\varepsilon) = \exp\{r/(K-r)\}.$$

This result is similar to the AEP lemma of [1]. It follows easily from Corollary 2 above since we have chosen N large enough so that $P(Z(K)) = 1 - o(1/n)$ (see the 'Sociability Lemma' of [1]). We use the inequality

$$\left(1 - \frac{r}{d+1}\right)^{-1} < \exp\{r/(d+1-r)\}.$$

Proof of Proposition 1. We first introduce a convenient method of generating our random graphs with ordered adjacency lists. Let a *configuration* L consist of a set L_0 of edges of G together with, for each vertex w of G, a linear order L_w on the vertices of G adjacent to w. Note that the number of such families of orders is

$$k = \prod_w \deg(w)!$$

where the product is over all vertices w of G, and where $\deg(w)$ denotes the degree of w in G. A configuration yields naturally a subgraph of G with ordered adjacency lists. We let each configuration L have probability $k^{-1}P\{R = L_0\}$, and this yields the random input we want.

For each set A of edges of G let $C(A)$ be the set of configurations L with $L_0 = A$ that yield the present position (that is, are such that the relevant initial segment of each adjacency list is as observed); and for $w = x, y$ let $C(A, w)$ be the set of configurations in $C(A)$ in which we find vertex w when we look at the next entry in the adjacency list for vertex v. Also, for each set A of edges let A' be obtained from A by swapping $\{v, x\}$ and $\{v, y\}$; that is, if we let $B = \{\{v, x\}, \{v, y\}\}$, then $A' = A \triangle B$ if $|A \cap B| = 1$ and $A' = A$ otherwise.

The key step in our proof of Proposition 1 is to establish the following combinatorial identity. Fix a set A of edges with $\{v, y\} \in A$, $\{v, x\} \notin A$, $D_x = d_x$, $D_y = d_y$. Then

$$\left(1 - \frac{r_x}{d_x + 1}\right) |C(A, y)| = \left(1 - \frac{r_y}{d_y}\right) |C(A', x)|. \tag{1}$$

In order to establish this claim let us first make one simple observation. Let $1 \leq r \leq s \leq t$, let T be a set of size t, let $S \subseteq T$ be of size s, and let z be an r-tuple of distinct elements in S: then the number of linear orders on T such that the first r elements in S form z equals $t!(s-r)!/(s!)$.

Let $D(A, y, x)$ be the set of configurations L in $C(A, y)$ in which the vertex v occurs in L_x after the last entry yet observed. Let L be any fixed configuration in $C(A, y)$, and let M denote the set of configurations in $C(A, y)$ which differ from L in at most the order L_x. Then by the above observation

$$|M| = (\deg(x)!)(d_x - r_x)!/d_x!$$

and

$$|M \cap D(A, y, x)| = (\deg(x)!)(d_x + 1 - r_x)!/(d_x+1)!$$
$$= |M|(1 - r_x/(d_x+1)).$$

Hence

$$|D(A, y, x)| = |C(A, y)|\left(1 - \frac{r_x}{d_x+1}\right). \tag{2}$$

Now define $D(A', x, y)$ to be the set of configurations L in $C(A', x)$ in which vertex υ occurs in L_y after the last entry yet observed. Then the above argument shows that

$$|D(A', x, y)| = |C(A', x)|\left(1 - \frac{r_y}{d_y}\right). \tag{3}$$

Finally, given any configuration L define a new configuration L' by swapping $\{\upsilon, x\}$ and $\{\upsilon, y\}$ in L_0 (that is, replacing L_0 by L_0') and swapping x and y in L_υ. The mapping $L \to L'$ yields a bijection between $D(A, y, x)$ and $D(A', x, y)$, and this together with (2), (3) completes the proof of the claim (1).

For integers $d_x' \geq r_x$ and $d_y' \geq r_y$, $d_y' \geq 1$ let

$$t(d_x', d_y') = \left(1 - \frac{r_y}{d_y'}\right) \Big/ \left(1 - \frac{r_x}{d_x'+1}\right).$$

Note that t is a decreasing function of d_x' and an increasing function of d_y'. Now if $\{\upsilon, x\} \in A$ and $\{\upsilon, y\} \in A$, then $A' = A$ and $|C(A, y)| = |C(A, x)|$. Hence using (1) and summing over those sets A of edges such that $D_x = d_x'$ and $D_y = d_y'$ we obtain

$$\begin{aligned}
&P\{\text{find } y, D_x = d_x', D_y = d_y'\}\\
&\quad = P\{\text{find } x, D_x = d_x', D_y = d_y', \{\upsilon, x\} \in R, \{\upsilon, y\} \in R\}\\
&\qquad + P\{\text{find } x, D_x = d_x' + 1, D_y = d_y' - 1, \{\upsilon, x\} \in R, \{\upsilon, y\} \notin R\}\, t(d_x', d_y').
\end{aligned}$$

[With the more general assumption about the random graph R discussed in Remark (ii) above this result holds as an inequality $\leq$, since if $C(A, y) \neq \emptyset$, then $P\{R = A\} \leq P\{R = A'\}$.] If we now sum over appropriate integers d_x', d_y' we obtain

$$\begin{aligned}
&P\{\text{find } y, D_x \geq d_x, D_y \leq d_y\}\\
&\quad \leq P\{\text{find } x, D_x \geq d_x, D_y \leq d_y, \{\upsilon, x\} \in R, \{\upsilon, y\} \in R\}\\
&\qquad + P\{\text{find } x, D_x \geq d_x + 1, D_y \leq d_y - 1, \{\upsilon, x\} \in R, \{\upsilon, y\} \notin R\}\, t(d_x, d_y)\\
&\quad \leq P\{\text{find } x, D_x \geq d_x, D_y \leq d_y\} \max\{1, t(d_x, d_y)\}
\end{aligned}$$

and Proposition 1 follows.

Note. If we wish only to establish Corollary 3, then it is sufficient to show that for any set A of edges

$$|C(A, y)| \leq |C(A', x)|;$$

and since we assume that $r_x = 0$ we always have $C(A, y) = D(A, y, x)$, and so the mapping $L \to L'$ yields an appropriate injection.

Acknowledgements

I would like to acknowledge helpful comments from Richard Karp and from the referee.

References

[1] D. Angluin and L.G. Valiant, Fast probabilistic algorithms for Hamiltonian paths, J. Comput. System Sci. 18 (1979) 155–193.

[2] R.M. Karp and R.E. Tarjan, Linear expected time algorithms for connectivity problems, J. Algorithms 1 (1980) 374–393.